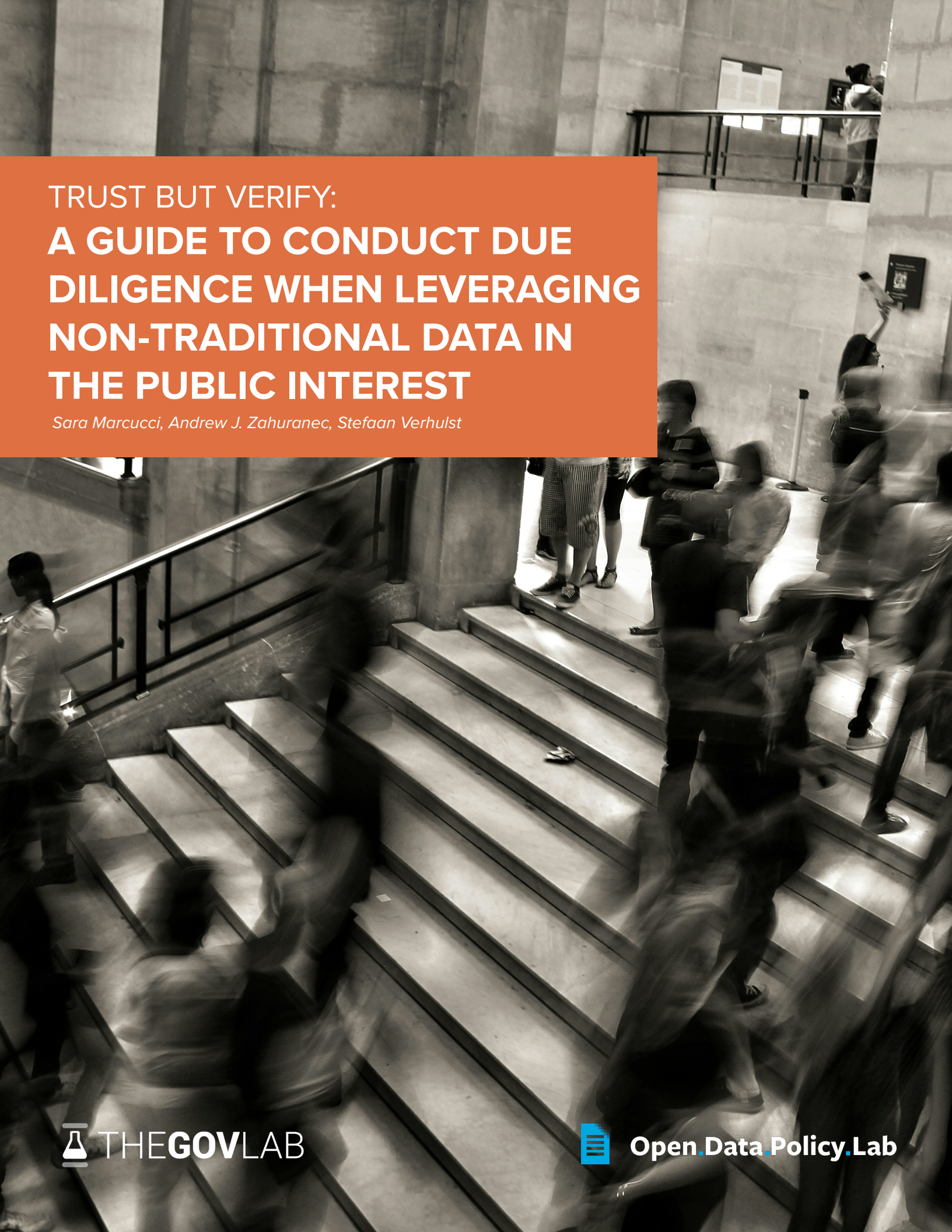


TRUST BUT VERIFY:

A GUIDE TO CONDUCT DUE DILIGENCE WHEN LEVERAGING NON-TRADITIONAL DATA IN THE PUBLIC INTEREST

Sara Marcucci, Andrew J. Zahuranec, Stefaan Verhulst

THEGOVLAB

Open.Data.Policy.Lab

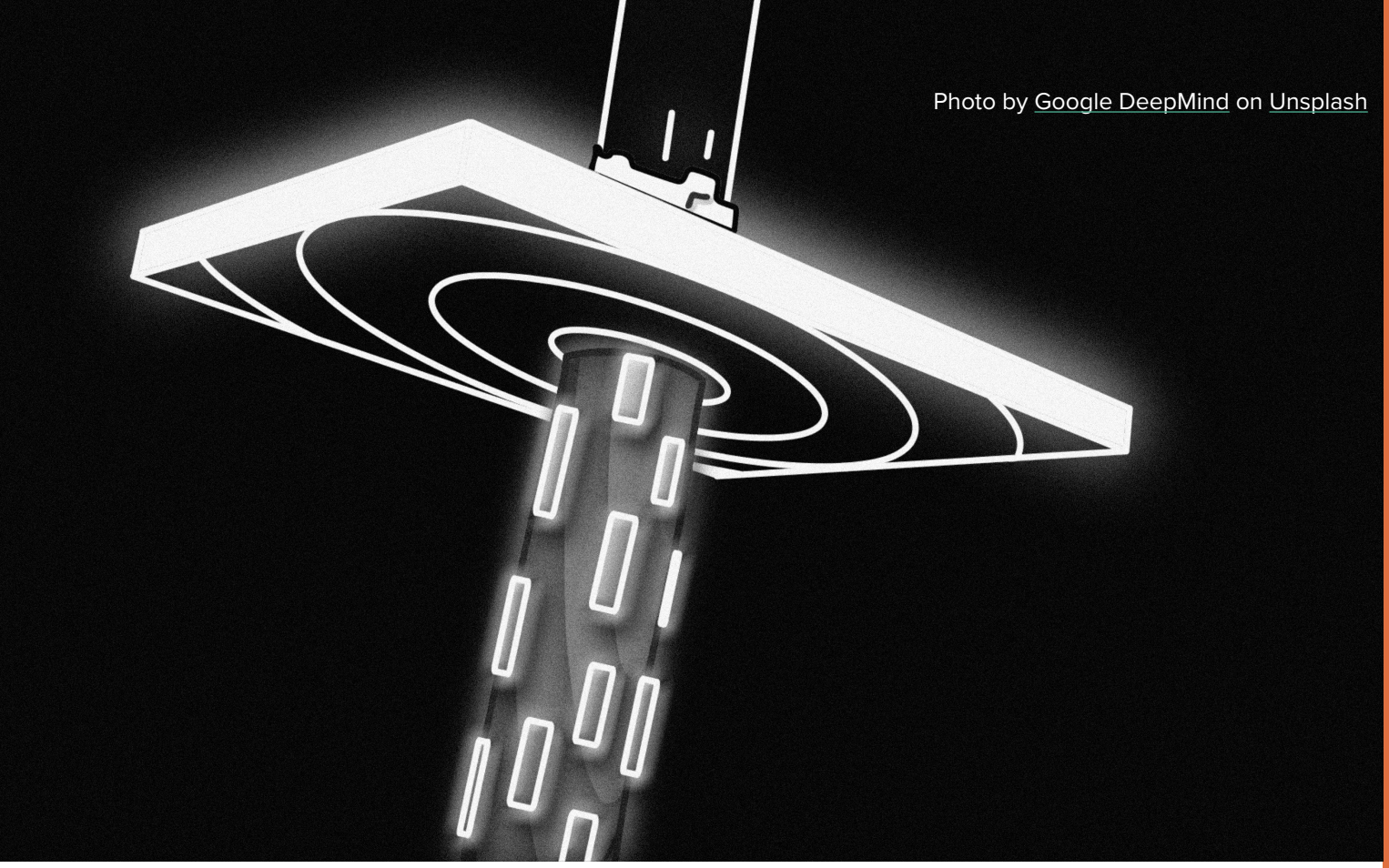


Photo by Google DeepMind on Unsplash

1. INTRODUCTION

In the evolving landscape of data-driven initiatives, due diligence is crucial for establishing responsible and effective data collaboration, especially when non-traditional data sources like satellite imagery, social media, or mobile data are involved. These types of data offer powerful insights but also present unique risks, especially when collaborations occur across unfamiliar organizations or in complex, uncharted contexts. The need for careful assessment extends to both sides of the partnership. On the one hand, it is important for data providers to feel confident that the recipient will handle their data responsibly to ensure that shared data will not be subject to misuse. On the other hand, it is crucial for data users to equally trust the integrity and reliability of the provider to ensure that the data they rely on is accurate, relevant, and compliant with established standards. This mutual trust is essential to building effective and secure data collaboration for public interest, where both parties feel confident in the quality, ethical handling, and shared value of the data.

Hesitancy toward engaging in data initiatives with a specific organization or public sector actor often stems from the risks related to unfamiliar or unknown data governance practices, particularly with non-traditional data, which may lack standardized protections and established protocols. Data initiatives involving different actors can reveal disparities in understanding and prioritizing data responsibility from both the provider's and recipient's

Cover Photo by José Martín Ramírez Carrasco on Unsplash

perspectives. These differing approaches can create challenges in maintaining a consistent, responsible approach to data governance in collaborative settings.

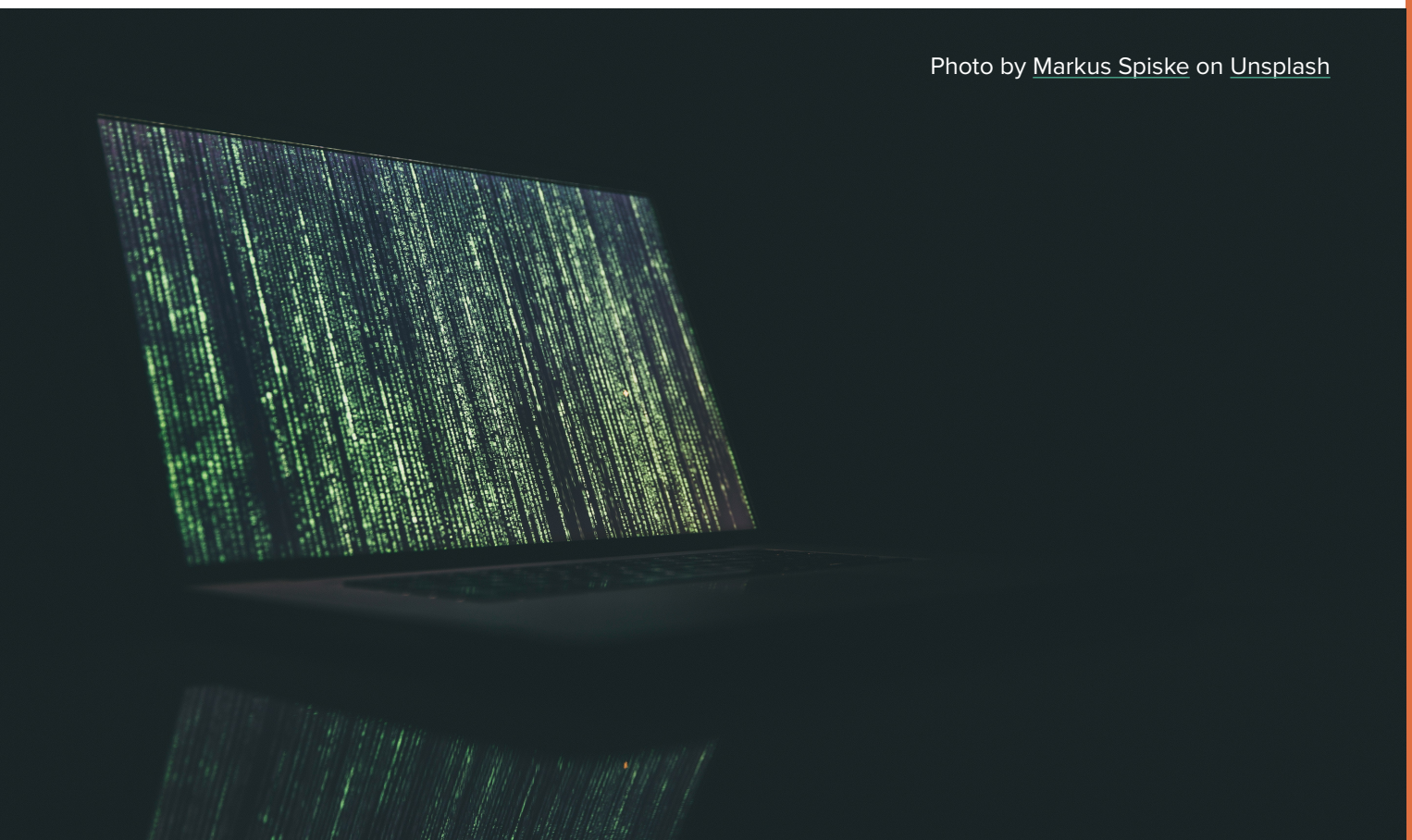
The intricacies and variety of versions of the data lifecycle further complicate the collaboration. Indeed, understanding how a potential partner will collect, process, analyze, share, use, and eventually dispose of the data introduces an additional layer of uncertainty. As a consequence, both data users and providers can face the challenge of comprehending these processes and evaluating potential risks to ensure responsible data practices throughout the lifecycle.

Cross-border collaborations exacerbate the challenge, with varying data protection regulations and cultural norms across different countries. Navigating these differences raises questions about legal compliance and other risks associated with international data sharing and collaboration.

The process of due diligence serves as a safeguard against potential risks and uncertainties that arise when engaging in non-traditional data reuse initiatives with external partners. It involves conducting thorough research, assessments, and evaluations to gain insights into the data recipient's capabilities, track record, and commitment to data responsibility.

This report provides a general guide for conducting such due diligence. It is intended to guide private and public organizations in non-traditional data initiatives that involve engagements with contexts and countries that are unfamiliar or potentially delicate from a data responsibility perspective.

Photo by Markus Spiske on Unsplash



2. VALUE PROPOSITION OF NON-TRADITIONAL DATA REUSE

1. **Expanding Insights Beyond Traditional Boundaries**: Non-traditional data sources, such as mobile data, satellite imagery, and social media analytics, provide unique insights that are often unattainable through conventional datasets. These sources enable organizations to explore new dimensions of understanding, capturing dynamic behaviors, preferences, and trends that enhance the richness of data-driven decision-making.

2. **Achieving Real-Time Responsiveness**: Many non-traditional data sources offer real-time or near-real-time insights, allowing organizations to adapt quickly to changing circumstances, whether in market trends, consumer behavior, or external conditions. Real-time responsiveness is invaluable for industries like retail, logistics, and finance, where timely information can drive competitive advantage and optimize strategic responses.

3. **Improving Decision-Making in Data-Sparse Environments**: In regions or sectors with limited access to traditional data sources, non-traditional data can fill critical gaps, providing contextually relevant information. This is especially beneficial in new or unfamiliar environments, where non-traditional data helps organizations understand local needs, preferences, and market dynamics, enabling more informed, culturally relevant decisions.

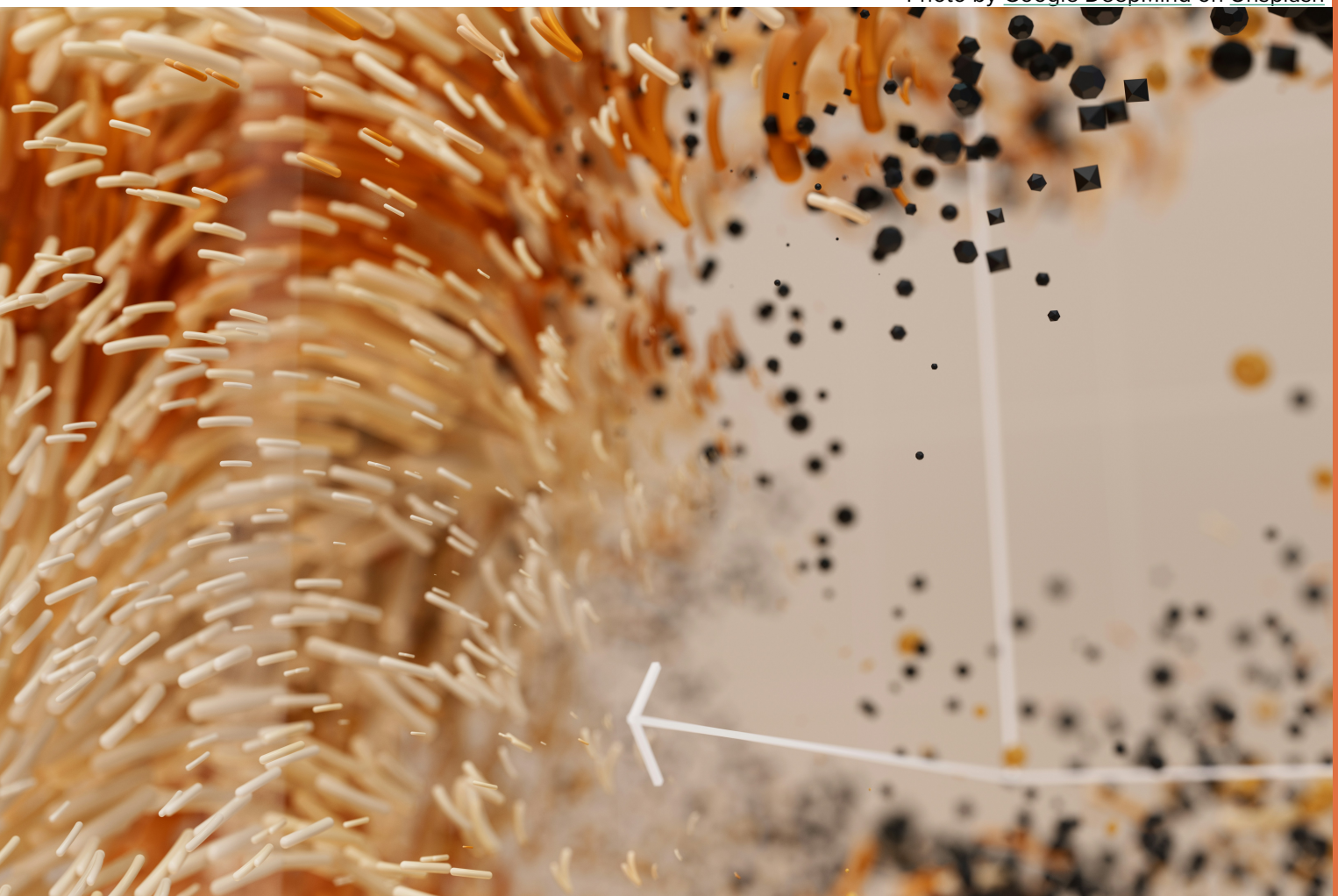
4. **Validating and Enriching Conventional Data**: Non-traditional data serves as a powerful tool to validate and complement traditional datasets, adding additional layers of context and accuracy. By cross-referencing with non-traditional sources, organizations can improve the reliability of their insights and reduce the risk of bias, helping to ensure that decisions are grounded in a fuller, more accurate picture.

5. **Enabling Targeted Resource Allocation**: Non-traditional data analytics allow organizations to identify specific needs or opportunities with a high degree of precision, supporting efficient resource allocation. This targeted approach is particularly valuable in unfamiliar settings, where non-traditional data can help pinpoint high-impact areas for investment, development, or service provision, thereby optimizing resource use.

6. **Navigating Complex, Cross-Border Contexts**: As organizations increasingly operate across borders, non-traditional data sources can help bridge the gaps posed by varying regulatory environments, data availability, and cultural norms.

By using responsibly sourced non-traditional data, organizations can adapt their strategies to diverse contexts, ensuring compliance and relevancy while respecting local contexts.

Photo by Google DeepMind on Unsplash



3. CRITERIA TO DETERMINE CONTEXTUAL RELEVANCE AND APPROPRIATENESS OF NON-TRADITIONAL DATA REUSE

Determining when non-traditional data is appropriate and can be leveraged responsibly involves carefully considering various factors. Criteria to help assess the appropriateness of non-traditional data can include:

1. **Relevance to the Problem:** Ensure that the non-traditional data is directly relevant to the issue or challenge you are addressing. It should provide meaningful insights or solutions.
2. **Data Availability:** Assess the availability of non-traditional data sources and map relevant organizations that might provide this data.
3. **Data Quality and Reliability:** Assess the quality and reliability of the data source. Is it accurate, up-to-date, and from a credible and trustworthy origin? Poor-quality data can lead to erroneous conclusions.
4. **Individual Consent and Group Agency:** Ensure that data is collected and used with the informed consent of individuals, while also establishing a social license to gain community acceptance and foster digital self-determination. Respect privacy and self/group-determination rights, especially for vulnerable populations.
5. **Data Security:** Implement robust security measures to safeguard the data against unauthorized access, breaches, or misuse.
6. **Ethical Considerations:** Consider the ethical implications of data collection and usage, especially when it involves vulnerable populations or sensitive personal information. Adhere to ethical standards and relevant frameworks.
7. **Bias Mitigation:** Evaluate the potential for bias in the data and employ strategies to identify and mitigate bias to ensure fairness and equity.
8. **Legal and Regulatory Compliance:** Ensure compliance with applicable laws and regulations, including data protection laws and regulations governing human rights.

RELEVANCE TO THE PROBLEM

Before exploring non-traditional data sources, it is crucial to articulate a specific question or problem that requires attention. This question serves as a **guiding principle, enabling the evaluation of whether non-traditional data can effectively address the issue at hand**. By

defining the problem, the context for data collection is set, and it becomes easier to determine the minimum viable data needed to find solutions. This approach minimizes the risk of data misuse and data overload and streamlines the selection of pertinent sources. Actions that can be taken to pursue this criteria can include:

- Clearly define the specific question or problem that needs attention, ensuring it's specific and actionable.
- Assess if the identified question aligns directly with the issue at hand, avoiding ambiguity.
- Outline the minimum data requirements necessary to address the problem, preventing unnecessary data collection.
- Establish clear criteria for selecting non-traditional data sources based on their direct relevance to the defined problem.

DATA AVAILABILITY

After establishing the question and relevance of non-traditional data to the problem, the next step is to assess the **availability of non-traditional data sources** capable of addressing it. This involves identifying potential organizations or entities that may possess the required data. These sources can span governmental agencies, private companies, research institutions, or social media platforms. The goal is to pinpoint sources with relevant information to inform the subsequent stages of analysis. Actions that can be taken to pursue this criteria can include:

- Identify specific organizations, platforms, or entities that may possess the required non-traditional data.
- Investigate the accessibility and availability of data from identified sources.
- Explore partnerships with governmental agencies, private companies, research institutions, and social media platforms.

DATA QUALITY AND RELIABILITY

Once relevant non-traditional data have been identified as available, it is important to assess their quality and reliability. Ensuring that the data is **accurate, up-to-date, and sourced from trustworthy origins** is essential. Poor-quality data can lead to misguided conclusions or decisions. To verify credibility, an assessment of the provenance and reputation of the data source is necessary, promoting confidence in the subsequent analytical processes. Actions that can be taken to pursue this criteria can include:

- Establish criteria for data quality and reliability, with an emphasis on accuracy and trustworthiness.
- Conduct assessments of the provenance and reliability of the data source to build confidence in subsequent analytical processes.

INDIVIDUAL CONSENT AND COMMUNITY AGENCY

Particularly when dealing with sensitive or personal information, adherence to informed consent principles is essential throughout the data lifecycle. This is even more crucial when dealing with vulnerable populations, where obtaining **explicit and meaningful consent** helps respect their autonomy and agency. To complement individual consent, establishing a **social license**—gaining community acceptance through ongoing engagement—can help ensure transparency, trust, and alignment with local values. A social license approach goes beyond traditional consent by securing broad community support and reducing the risks of opposition or non-compliance, thereby increasing the likelihood of sustained collaboration and effective data reuse outcomes. This approach can also empower groups to determine how their information is used in a way that aligns with their interests, preferences, and expectations. Actions that can be taken to pursue this criteria can include:

- Develop explicit and accessible consent processes for data subjects.
- Engage community stakeholders throughout the data lifecycle.
- Incorporate community feedback continuously to ensure data use aligns with collective expectations and local values.
- Document consent and engagement procedures, particularly when working with vulnerable groups.

DATA SECURITY

To safeguard against potential threats, robust security measures must be implemented throughout the data lifecycle. This includes **encryption, stringent access controls, and compliance with relevant data protection regulations**. The objective is to fortify the data against unauthorized access, breaches, or any form of misuse, thereby maintaining the confidentiality and integrity of the collected information. Actions that can be taken to pursue this criteria can include:

- Implement encryption and stringent access controls to secure non-traditional data.
- Ensure compliance with data protection regulations and industry best practices.
- Develop a detailed data security strategy aligned with the specific needs of the project.
- Regularly audit and update security measures to respond to evolving threats and vulnerabilities.

ETHICAL CONSIDERATIONS

It is vital to approach the use of non-traditional data sources with a strong ethical foundation. **Meaningful individual and group consent/assent, data minimization, and transparency** in data handling are some of the principles that can be prioritized. This involves a conscientious approach throughout the data lifecycle, ensuring that practices align with **broader ethical**

principles and do not compromise the rights or well-being of groups and individuals. Actions that can be taken to pursue this criteria can include:

- Develop a process to handle requests to share or receive data. The team might, for example:
    - Assess the purpose of sharing and its legitimacy;
    - Evaluate what is necessary to share (e.g. raw data, clean data, personal data) and the risk of each;
    - Define what data is sensitive in an operational context; and
    - Identify what data sharing arrangements are necessary to ensure that partners uphold data responsibility principles and ethical requirements.
- Within the multi-functional team, assess what is needed to secure the data to guarantee the privacy of interested data subjects and prevent unwanted exposure. Practitioners might implement:
    - Physical security (e.g. making data only accessible from certain computers or office locations);
    - Technological security (e.g. making it so that certain user profiles have restricted access; and
    - Procedural (e.g. having standard operating procedures with an approval chain to create users, exchange data).

BIAS MITIGATION

Acknowledging the potential for bias in data is a critical aspect of responsible data analysis. It is indeed important to **assess whether the collection and use of non-traditional data inadvertently perpetuate biases or contribute to unfair treatment of data subjects**, especially towards vulnerable populations. Evaluating and addressing existing and potential biases, whether inherent or introduced during the data and decision cycles, is imperative for establishing whether or not the use of non-traditional data can be responsible. Actions that can be taken to pursue this criteria can include:

- Identify potential biases throughout the non-traditional data lifecycle;
- Ensure that data has been cleaned before it is analyzed, as data that is not well organized before analysis will not generate reliable findings;
- Evaluate the type of bias and develop protocols to address biases specific to vulnerable populations.
- Ask critical questions to minimize biases and ensuring fairness and equity, such as:
    - What categories are relevant to the decisions that we hope to make with the data?
    - What disaggregation points are relevant to the purpose (e.g. location, country/ region of origin);
    - What are the limitations of the dataset and what can it *not* represent?

LEGAL AND REGULATORY COMPLIANCE

Legal and regulatory compliance is a fundamental aspect of any data initiative, including non-traditional data activities. This encompasses **adherence to established laws, regulations, and industry standards** governing the collection, processing, storage, sharing and use of data. It serves as a safeguard against legal ramifications and can reinforce ethical practices in handling non-traditional data. Actions that can be taken to pursue this criteria can include:

- Ensure adherence to specific data protection laws and regulations applicable to the project.
- Regularly review compliance measures to stay updated with evolving legal requirements.
- Establish protocols for handling data in accordance with local and international laws.
- Train personnel to understand and comply with the legal landscape governing non-traditional data use.

Photo by Alan Fung on Unsplash

4. DUE DILIGENCE FOR ENSURING RESPONSIBLE DATA PRACTICES WHEN LEVERAGING NON-TRADITIONAL DATA FOR THE PUBLIC INTEREST

Due diligence is a continuous process, and is to be revisited on a regular basis. Indeed, because of the vast diversity in contexts, organizations, and technologies, no two due diligence processes will be the same or stay the same. The process illustrated here is indeed only meant to be a starting point and, in order to be effective, needs to be tailored to the specific legal, cultural and political contexts and needs of the organization.

In particular, we present six steps:
(1) Determination of Due Diligence Scope,
(2) Internal Data Collection,
(3) Risk ranking and Red Flag identification,
(4) Additional Due Diligence Tool for High-Risk Cases,
(5) Approval Based on Risk Level, and
(6) Post Engagement Due Diligence.

STEP 1: DETERMINATION OF DUE DILIGENCE SCOPE

WHAT: The first step in the due diligence process is to define the scope of the assessment. Organizations are to identify the data subjects and sources, partners, and contexts they will be engaging with to understand the potential risks involved. This step is essential as it lays the foundation for the rest of the process, allowing organizations to focus their efforts on the areas that matter most.

WHO: The due diligence team (Figure 1).

Example: Due Diligence team		
Operations	**Legal and Policy**	**Domain Experts**
Partnerships: Purpose of data use and details of data request, general partner relations	**Legal:** Laws & regulations that govern data collection, use, and storage in country jurisdiction. Manage contracts.	**Data Owner:** Ensure data fit for purpose and use. Manage enquiries on nature and content of data set.
Project/Delivery Management: Timeline and deliverable management	**Data policy analyst:** Enablement of data to be shared responsibly by ensuring internal policies are followed.	**Country of Origin expert:** Support Legal and Partnerships with socio political and economic insights.
	Information Assurance: Manage risks related to the use, processing, storage, and transmission of data.	**Subject matter expert:** Support Legal, Partnerships and data owner with subject matter insights.

Figure 1: Example team of specialists, and their roles, who could make-up a due diligence team. We have avoided using specific job titles as these are not standardized across geography, sector or industry. We have opted to rather name teams or departments that should be involved.

HOW: Conducting desk research and developing a checklist to identify risks. More specifically, it may prove useful to identify the risks across the different stages of the data lifecycle (Figure 2).

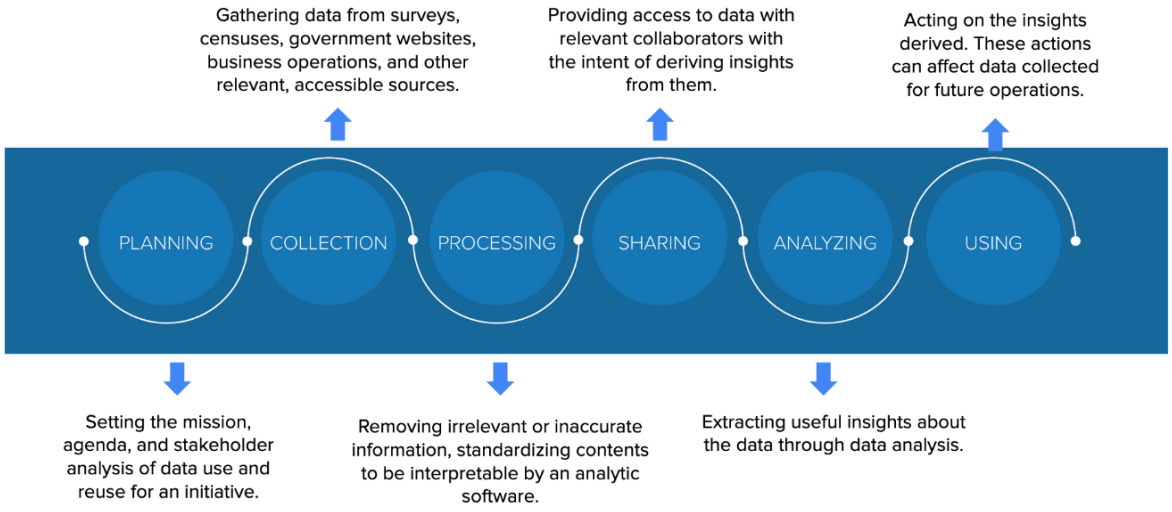


Figure 2: The Data Lifecycle by The GovLab.

Examining risks at different stages of this journey may enable organizations to gain a comprehensive understanding of potential vulnerabilities and threats associated with their data initiatives.

Risks across the data lifecycle may include:

Stage	Risks
Planning	Unclear purpose and goals of the project
	Financial resources and stakeholder partnerships hard to find and maintain
Collecting	Data subjects and sources and their consent not being protected
	Poor quality (duplication or inconsistencies) or dirty data, data bias/non-representation
Processing	Poor anonymization and privacy of data subjects and sources
	Insufficient security provisions, inaccessibility, aggregation and correlations challenges
Sharing	Lack of trust and communication among partners, conflicting jurisdictions, different levels of security
	Poor balance between interoperability and context-rich, relevant data
Analyzing	Poor focus and rigor in the data analysis, inaccurate data modeling, poor problem and definition design
	Black boxes, blindspots, unexplainable automated decision making, inequity production and reproduction (eg. biased algorithms)
Using	Faulty reporting (eg. misinterpretation), malicious actors using findings for unforeseen and/or harmful purposes
	Poor communication strategy and external use and repurpose of findings

Table 1: The Risks Across the Data Life Cycle Framework

STEP 2: INTERNAL DATA COLLECTION

WHAT: During Step 2 of the due diligence process, the focus shifts towards internal data collection, a pivotal phase in understanding the contextual and geographical risk factors surrounding the data initiative. This step aims to gather critical information about the regulatory, legal, and ethical landscape in which the organization will operate. In particular, this step may involve collecting information about data protection regulations, rule of law and due process, corruption tolerance, and human rights records.

WHO: The due diligence team, consisting of researchers, project managers, and program heads, takes the lead in this phase. Working in tandem with a trusted law firm, the team collaboratively explores the regulatory and legal frameworks that govern data practices in the specific context and geographic location of the data initiative.

HOW: The due diligence team may conduct desk research and consult various sources, such as government publications, legal databases, reputable research papers, and reports from non-governmental organizations. The involvement of legal experts from the due diligence team can provide valuable insights into the regulatory landscape and help identify potential risks and compliance requirements. Furthermore, legal experts may review pertinent documents, such as data protection policies, legislation, and human rights reports, to assess the potential impact of the data initiative on individual rights and the overall legal landscape. Additionally, the due diligence team may engage in contextual interviews with local experts or stakeholders to gain a deeper understanding of the practical implications of the regulatory environment and cultural norms. Based on the data collected, the due diligence team may develop a checklist to systematically assess and determine the due diligence scope, ensuring that all relevant risk factors are carefully considered and evaluated.

STEP 3: RISK RANKING AND RED FLAG IDENTIFICATION

WHAT: In step 3, organizations assess the public benefits and risks associated with data initiatives in the identified context. This involves ranking potential risks based on severity and identifying any red flags that require immediate attention. The goal is to gain a comprehensive understanding of the risks to make informed decisions moving forward.

WHO: In Step 3, similarly to Step 2, the due diligence team, in collaboration with a trusted law firm, takes the lead. Together, they work diligently to assess the potential public benefits of the data initiative while also identifying and ranking the risks associated with its implementation.

HOW: To achieve this, the due diligence team may consider creating and using a risk ranking framework that provides a structured and standardized guide for analyzing the severity and significance of each identified risk. This will need to be developed according to the specific context and needs of the organization.

STEP 4: ADDITIONAL DUE DILIGENCE TOOL FOR HIGH-RISK CASES

WHAT: In cases where high-risk factors are identified during the initial due diligence process, additional external due diligence becomes necessary. To this end, we have developed a Due Diligence Diagnostic Tool that illustrates the critical aspects and questions that a Committee of Experts may consider when faced with such circumstances. The tool is organized according to the stages of the data lifecycle (Figure 1), and delves into a thorough investigation to determine whether the engagement is still viable through mid-way solutions or if it should be reevaluated altogether.

STEP 5: APPROVAL BASED ON RISK LEVEL

WHAT: In the final stages of the due diligence process, Step 5 involves seeking approval from the Committee of Experts (Figure 3). The goal is to confirm that the identified risks have been adequately mitigated or reduced to a level that allows the data project to proceed without further mitigation measures.

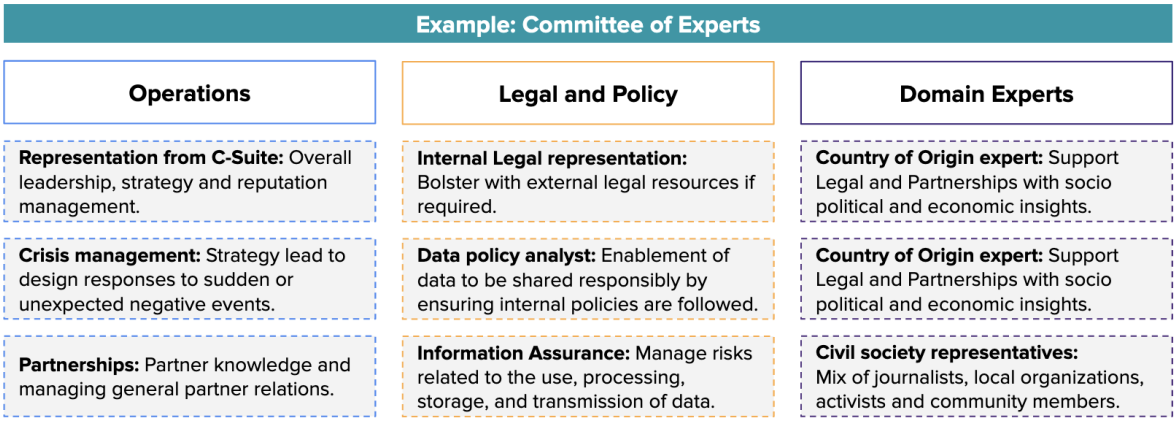


Figure 3: Example team of specialists, and their roles, who could make-up a Committee of Experts. We have avoided using specific job titles as these are not standardized across geography, sector or industry.

WHO: The Committee of Experts plays a pivotal role in this step. They are responsible for reviewing the due diligence findings and providing their expertise in assessing the adequacy of risk mitigation efforts.

HOW: To facilitate the approval process, the due diligence team develops a checklist using insights from the Risks Across the Data Life Cycle framework, adopted in Step 1. The Committee of Experts will use this checklist to confirm whether the risks identified in Step 4 have been mitigated or not.

STEP 6: POST ENGAGEMENT DUE DILIGENCE

WHAT: Step 6 involves the process of post-engagement due diligence monitoring. After the data initiative has been initiated and the organization has engaged with the partnering organization or government, this phase focuses on continuous assessment and monitoring to ensure ongoing responsible data practices. Due diligence is not a one-time process; it requires ongoing assessments to adapt to changing contexts, regulations, and organizational needs. Regular reviews help organizations stay accountable and responsive to evolving data risks.

WHO: The responsibility for post-engagement due diligence monitoring may lie with the head of programs, supported by the due diligence team and internal lawyers. Additionally, involved data subjects, whose data is being used, are considered key stakeholders and may provide valuable insights during the monitoring process.

HOW: To conduct post-engagement due diligence monitoring effectively, the following measures may be implemented: Establishing Monitoring Protocols: Clear and robust monitoring protocols are put in place to regularly assess the data initiative's progress and adherence to ethical standards and regulations. These protocols outline the scope, frequency, and methods of monitoring.

1. **Establishing Monitoring Protocols:** Clear and robust monitoring protocols are put in place to regularly assess the data initiative's progress and adherence to ethical standards and regulations. These protocols outline the scope, frequency, and methods of monitoring.
2. **Regular Audits:** Scheduled audits are conducted to review the data initiative's activities, ensuring that the initial risk assessment and compliance requirements continue to be met. These audits help identify any new risks or ethical concerns that may have emerged during the engagement.
3. **Data Subjects Engagement:** Ongoing engagement with data subjects is maintained to gather feedback and assess their experiences with the data initiative. This direct interaction helps gauge the impact on data subjects and identify any potential privacy or ethical issues.
4. **Responsive Actions:** If any new risks or ethical challenges are identified during post-engagement monitoring, the due diligence team, in collaboration with internal lawyers, takes prompt and responsive action to address and mitigate these issues. Proactive measures are essential to maintain data responsibility throughout the data initiative.
5. **Periodic Reporting:** Regular reporting on post-engagement due diligence findings is shared with relevant stakeholders, including organizational leaders, data subjects, and partnering organizations or governments. Transparent reporting ensures accountability and builds trust among stakeholders.

Photo by Ryoji Iwata on Unsplash

5. CONCLUSION

This paper aimed to provide a general guide that can help organizations navigate uncharted waters when partnering with different entities in unfamiliar contexts. Throughout this process, we have emphasized the significance of conducting due diligence to enable responsible data use and reuse. It is essential to acknowledge that due diligence is not a one-size-fits-all process; the process demands constant adaptation to the specific legal, cultural, and political landscapes in which organizations operate. The steps outlined here serve as a starting point, offering a roadmap for private and public organizations to assess the risks associated with handling data from diverse sources, collaborating with various partners, and operating in different contexts.

APPENDIX A: ADDITIONAL DUE DILIGENCE AND DIAGNOSIS - WHAT TO CONSIDER IN CASE HIGH RISK IS IDENTIFIED FOR A PARTICULAR PROJECT OR DATASET

The essential goal of due diligence is to assess the potential risks of using private sector non-traditional data. Additional due diligence may be necessary when potential risks are identified, to ultimately find potential solutions or–alternatively–cancel the engagement.

This tool illustrates the additional due diligence and diagnosis to be conducted **in case high risk is identified** during the due diligence process, particularly unfolding Step 4: Additional External Due Diligence. The aim is to show the issues and questions the project team could consider in case of high risk, when trying to determine whether the engagement is still worth establishing through **mid-way solutions** or not.

1. Questions to be asked for each risk across the data lifecycle

FRAMEWORK: The mitigation strategies are very broad and only serve as a starting point for the project team to consider different elements and options in case of identified high risk. These will need to be tailored to each case and further refinement of such strategies will have to take place once the specific context is identified.

Stage	Risks identified	Additional Due Diligence and Diagnosis: What could the project team consider for each risk?
Planning	The project has unclear purposes and goals	All actors involved in the project may consider completing a project case if one has not been provided. At a minimum, the need and purpose of the project have to be defined and agreed on. POTENTIAL QUESTIONS TO ASK: • Why does the project have an unclear purpose and undefined goals? • Who is responsible for defining these? • Are there established metrics and indicators of success? If not, can we establish these?

	Financial resources and stakeholder partnerships are difficult to find and maintain	It's recommended that all actors involved in the project come together and analyze the reasons for a lack of resources and partnerships. The most desired outcome is for clear processes and accountabilities to be established.
		
		POTENTIAL QUESTIONS TO ASK:
		• What financial resources are missing?
		  ○ Why are they missing?
		  ○ Who and what is needed to rectify the situation?
		• Which stakeholder partnerships are needed?
		  ○ Why are these relationships difficult to establish and/or maintain?
		  ○ What can be done to improve the current situation? More collaboration? Greater transparency?

Collecting	The consent of data subjects and sources not being protected	Identify the reasons for this and then determine the next appropriate steps to take. As a guide: • If the cause has to do with national or local policy and legal drivers, the engagement may have to be canceled. • If the cause has to do with poor management, but there is indeed willingness to uphold consent, responsible actors are to be identified and made accountable. • If the cause has to do with negligence, it should be clear whether the actors involved are indeed interested in protecting data subjects' consent, and if and how it may be feasible to do so.

	Poor quality (duplication or inconsistencies) or dirty data, data bias/non-representation	The actors involved in the collection of the data are to be involved in the mitigation strategy, which is to be co-designed with them.
		• If data is duplicated or dirty, delete double entries and clean the data.
		• If data is biased, further investigations need to be carried out to understand why.
		
		POTENTIAL QUESTIONS TO ASK:
		• Who is responsible for cleaning the data?
		  ○ Why is there poor quality data? How can we ensure we minimize the chances of this happening again in the future?
		• In what way is the data biased? E.g. Due to poor representation, discrimination against certain groups and perpetuation of stereotypes.
		  ○ Can this be addressed?
		  ○ Who should be involved in the process of improving the quality of the dataset so as to minimize the risks of unintended consequences ?
		  ○ How can we improve data quality assurance processes over time?

Processing	Poor anonymization and privacy of data subjects and sources	Anonymizing data ensures the risk of re-identification is minimized. If it is discovered that data has not been, or has been poorly anonymized, then you may wish to consider applying the following treatments to the dataset in order to safeguard privacy: • Suppression > Remove data from a dataset. Best applies to any direct identifiers. • Randomization > Add noise or shuffle values while maintaining patterns in the dataset. • Pseudonymization > While not a method of anonymization, this method can minimize the chances of data subjects and sources being reidentified. Techniques include encryption, tokenization and hashing. POTENTIAL QUESTIONS TO ASK: • Considering the nature of the dataset, which is the best anonymization technique to apply? Once anonymization techniques have been applied: • Is it possible to single out and identify an individual? • Can public records be linked together and used to identify an individual in the dataset?

	Insufficient security provisions, inaccessibility, aggregation and correlations challenges	All stakeholders could meet to discuss the severity and variety of harms that are possible should the data be released. Consider delaying the launch of the data/project until the risk is sufficiently reduced. QUESTIONS TO ASK: • Do we have the required resources to tackle the challenge and minimize risk to a satisfactory level? • What actions need to be taken to minimize the severity and variety of harms that could result from this data?
Sharing	Lack of trust and communication among partners, conflicting jurisdictions, different levels of security	It is important to understand the legal and ethical basis for sharing personal data. This will be geography and context-dependent. POTENTIAL QUESTIONS TO ASK: • What is fueling the distrust between stakeholders? What can be said and done to reduce fear and improve cooperation? • Do we understand the nature of the data and the reasons for sharing it? • Are we exercising the principle of data proportionality? • Are adequate security measures in place to keep data safe for as long as it is needed?
	Poor balance between interoperability and context-rich, relevant data	Successful sharing of data requires datasets to be interoperable across many systems. It also requires data to be meaningful and useful. POTENTIAL QUESTIONS TO ASK: • What needs to be done to improve the interoperability features of this dataset? • What context can be added to the data to improve its relevance? • What processes do we need to consider implementing during the collection phase to ensure that data is more meaningful in the future?
Analyzing	Poor focus and rigor in the data analysis, inaccurate data modeling, poor problem and definition design	If it is discovered that data has been poorly or incorrectly analyzed, then focus stakeholder attention on what can be improved before any data is shared. POTENTIAL QUESTIONS TO ASK: • What needs to be done to improve the quality of analysis of this dataset? • Do we have the required resources to tackle the challenge and minimize risk to a satisfactory level?

	Black boxes, blindspots, unexplainable automated decision making, inequity production and reproduction (eg. discriminatory algorithms)	Unexplainable systems have the potential to cause a number of harms. Consider delaying the launch of the data/project until the risk is sufficiently reduced. POTENTIAL QUESTIONS TO ASK: • Can we improve the explainability of this data system? • If not, is it possible to implement a system that can be explained?
Using	Faulty reporting (eg. misinterpretation), malicious actors using findings for unforeseen and/or harmful purposes	You may wish to employ the skills of an experienced PR or communications professional during a situation of this nature. POTENTIAL QUESTIONS TO ASK: • Has this potentially been a misunderstanding? Could we contact the actors and attempt to set the record straight? • If not, can we develop a fact-correcting communications strategy in order to rectify any disinformation and misinformation? • Who could we partner with to amplify these messages and ensure we get the facts in front of more people?
	Poor communication strategy and external use and repurpose of findings	Working with an experienced PR or communications professional during a situation of this nature is recommended. POTENTIAL QUESTIONS TO ASK: • What immediate and short-term tactics can we use to improve the public narrative around this work? • Can we lean on any partners or influential individuals to support us during this endeavor? • What is our long-term strategy to ensure we course correct over time and are still able to achieve our desired outcomes?